On monotone metric of classical channel and distribution spaces: asymptotic theory

Keiji Matsumoto

August 24, 2010

Abstract

The aim of the manuscript is to characterize monotone metric in the space of Markov map. Here, metric is not necessarily Riemanian, i.e., may not be the inner product of the vector with itself.

So far, there have been plenty of literatures on the metric in the space of probability distributions and quantum states. Among them, Cencov and Petz characterized all the monotone metrics in the classical and quantum state space. As for channels, however, only a little is known about its geometrical structures.

In that author's previous manuscript , the upper and the lower bound of monotone channel metric was derived using resource conversion theory, and it is proved that any monotone metric cannot be Riemanian.

Due to the latter result, we cannot rely on Cencov's theory, to build a geometric theory consistent across probability distributions and channels. To dispense with the assumption that a metric is Riemanian, we introduce some assumptions on asymptotic behavior, *weak asymptotic additivity* and *lower asymptotic continuity*. The proof utilizes resource conversion technique. In the end of the paper, an implication on quantum state metrics is discussed.

1 Introduction

The aim of the manuscript is to characterize monotone metric in the space of Markov map. Here, metric means the square of the norm defined on the tangent space, and not necessarily Riemanian, nor induced from an inner product.

So far, there have been plenty of literatures on the metric in the space of probability distributions and quantum states. Cencov, sometime in 1970s, proved the monotone Riemanian metric in probability distribution space is unique up to constant multiple, and identical to Fisher information metric [5]. He also

discussed invariant connections in the same space. Amari and others independently worked on the same objects, especially from differential geometrical view points, and applied to number of problems in mathematical statistics, learning theory, time series analysis, dynamical systems, control theory, and so on[1][2]. Quantum mechanical states are discussed in literatures such as [2][6][9][9][15]. Among them, Petz [15] characterized all the monotone Riemanian metrics in the quantum state space using operator mean theory.

As for channels, however, much less is known. To my knowledge, there had been no study about axiomatic characterization of distance measures in the classical or quantum channel space, except for the author's manuscript [11]. In that manuscript, the upper and the lower bound of monotone channel metric was derived using resource conversion theory, and it is proved that any monotone metric cannot be Riemanian.

The latter result has some impact on the axiomatic theory of the monotone metric in the space of classical and quantum states, since both Cencov [5] and Petz [15] assumed metrics are Riemanian. Since classical and quantum states can be viewed as channels with the constant output, it is preferable to dispense with this assumption. Recalling that the Fisher information is useful in asymptotic theory, it would be natural to introduce some assumptions on their asymptotic behavior. Hence, we introduced *weak asymptotic additivity* and *lower asymptotic continuity*. By these additional assumptions, we not only recovers uniqueness result of Cencov [5], but also proves uniqueness of the monotone metric in the channel space.

In this proof, again, we used resource conversion technique. A difference from usual resource conversion technique is that asymptotic continuity is replaced by a bit weaker *lower asymptotic continuity*. The reason is that the former condition is not satisfied by Fisher information.

In the end, there is an implication on quantum state metrics.

2 Notations and conventions

In discussing probability distributions, the underlying set is denoted by $\otimes$. In discussing channels, $\otimes_{\mathrm{in}}$ ($\otimes_{\mathrm{out}}$) denotes the totality of the inputs (outputs). In this paper, they are either $\{1, \cdots, k\}$ or $\mathbb{R}^d$. x,y, etc. denotes an element of $\otimes_{\mathrm{in}}$,$\otimes_{\mathrm{out}}$, $\otimes$. Also, $x^n = (x_1, x_2, \cdots, x_n)$, $y^n = (y_1, y_2, \cdots, y_n)$, etc. denotes an element of $\otimes^{\times n}$, $\otimes_{\mathrm{in}}^{\times n}$,$\otimes_{\mathrm{out}}^{\times n}$.

Random variable taking values in $\otimes$, $\otimes_{\mathrm{in}}$,$\otimes_{\mathrm{out}}$ are denoted by X, Y, while random variable taking values in $\otimes^{\times n}$, $\otimes_{\mathrm{in}}^{\times n}$,$\otimes_{\mathrm{out}}^{\times n}$ are denoted by X^n, Y^n. The dsitribution of X is denoted by $\mathbb{P}_X$, while its density (with respect to lebesgue measure or counting measure depending on the underlying set) is denoted by p_X. $\mathbb{P}_{X|Y}$ and $\mathrm{p}_{X|Y}$ denotes the onditional distribution and its density, respectively. In this paper, the existence of density with respect to a standard underlying measure μ (counting measure for $\{1, \cdots, k\}$, and Lebesgue measure for $\mathbb{R}^d$) is always assumed. Hence, by abusing the term, we sometimes say 'distribution p'. By $\mathcal{P}$, $\mathcal{P}_{\mathrm{in}}$, and $\mathcal{P}_{\mathrm{out}}$ we denote the totality of the probability density functions over $\otimes$, $\otimes_{\mathrm{in}}$, and $\otimes_{\mathrm{out}}$, respectively.

Channel Φ is a linear map from probability distributions over to $\otimes_{\mathrm{in}}$ to those over $\otimes_{\mathrm{out}}$, but also considered as a map from $L_1(\otimes_{\mathrm{in}})$ to $L_1(\otimes_{\mathrm{out}})$. Hence, we use notation such as $\Phi(\mathbb{P}_X)$, as well as $\Phi(\mathrm{p}_X)$. The totality of channels is denoted by $\mathcal{C}$. If there is a need to indicate input and output space, we use the notation such as $\mathcal{C}(\mathcal{P}_{\mathrm{in}}, \mathcal{P}_{\mathrm{out}})$. Φ^* denotes the dual map of Φ,

$$\int f(x)\,\Phi(\mathrm{p}_X)(x)\,\mathrm{d}\mu(x) = \int \Phi^*(f)(x)\,\mathrm{p}_X(x)\,\mathrm{d}\mu(x).$$

A tangent space is denoted by a notation $\mathcal{T}_{\cdot}(\cdot)$. δ, δ' etc. denotes an element of $\mathcal{T}_p(\mathcal{P})$ (the tangent space to the set $\mathcal{P}$ at the point p) etc, w hile Δ, Δ' etc denotes an element of $\mathcal{T}_\Phi(\mathcal{C})$ etc. In the paper, we identify $\delta \in \mathcal{T}_p(\mathcal{P})$ with an element of L^1 in the form of $c(p_1 - p_2)$, where $p_1,\, p_2 \in \mathcal{P}$. Hence, the differential map of Φ is also denoted by Φ, by abusing the notation. L is a random variable defined by

$$L(x) = \frac{\delta(x)}{p(x)},\quad x \in \Omega.$$

and its low is under p, unless otherwise mentioned. Also, Δ is identified with a linear map in the form of $c(\Psi_1 - \Psi_2)$, where $\Psi_1,\, \Psi_2 \in \mathcal{C}$.

A pair $\{p, \delta\}$ and $\{\Phi, \Delta\}$ is called *local data at* p and Φ, respectively, since it decides local behaviour of one-parameter family of distributions at the point p and Φ, respectively. We denote by $\mathrm{N}\left(a, \sigma^2\right)$ and $\delta\mathrm{N}\left(a, \sigma^2\right)$ the Gaussian distribution with mean a and variance σ^2 and singed measure defined by

$$\delta\mathrm{N}\left(a, \sigma^2\right)(B) := \frac{1}{\sqrt{2\pi}\sigma}\int_B \frac{x-a}{\sigma^2}\exp\left[-\frac{1}{2\sigma^2}(x-a)^2\right]\mathrm{d}x,$$

respectively. Thus, the local data $\left\{\mathrm{N}\left(a, \sigma^2\right), \delta\mathrm{N}\left(a, \sigma^2\right)\right\}$ describes local behaviour of Gaussian shift family $\left\{\mathrm{N}\left(\theta, \sigma^2\right)\right\}_{\theta\in\mathbb{R}}$ at $\theta = a$.

The symbol '$\otimes$' means direct product of vectors. Given $f \in L_1(\otimes_1)$ and $g \in L_1(\otimes_2)$, $f \otimes g$ is defined by

$$f \otimes g(x_1, x_2) = f(x_1)\, f(x_2).$$

The linear span of $\{f \otimes g\}$ is denoted by $L_1(\otimes_1) \otimes L_2(\otimes_2) (= L_1(\otimes_1 \times \Omega_2))$. Also, given $\Phi_1 \in \mathcal{C}(\mathcal{P}_{\mathrm{in},1}, \mathcal{P}_{\mathrm{out},1})$, $\Phi_2 \in \mathcal{C}(\mathcal{P}_{\mathrm{in},2}, \mathcal{P}_{\mathrm{out},2})$, $\Phi_1 \otimes \Phi_2 \in \mathcal{C}(\mathcal{P}_{\mathrm{in},1} \otimes \mathcal{P}_{\mathrm{in},2}, \mathcal{P}_{\mathrm{out},1} \otimes \mathcal{P}_{\mathrm{out},2})$ is defined by the relation

$$\Phi_1 \otimes \Phi_2(f \otimes g) = \Phi_1(f) \otimes \Phi_2(g)$$

and linearity. For a real valued random variable F_1 and F_2 over Ω_1 and Ω_2, respectively, $F_1 \otimes F_2$ is a random variable over $\Omega_1 \times \Omega_2$ with

$$F_1 \otimes F_2(x_1, x_2) = F(x_1)\, F(x_2).$$

We use abbraviations such as $f^{\otimes n} := f \otimes f \otimes \cdots \otimes f$, and

$$
\begin{aligned}
\delta^{(n)} &:= \delta \otimes p^{\otimes n-1} + p \otimes \delta \otimes p^{\otimes n-2} + \cdots + p^{\otimes n-1} \otimes \delta \in \mathcal{T}_p\left(\mathcal{P}^{\otimes n}\right), \\
L^{(n)} &:= L \otimes 1^{\otimes n-1} + 1 \otimes L \otimes 1^{\otimes n-2} + \cdots + 1^{\otimes n-1} \otimes L, \\
\Delta^{(n)} &:= \Delta \otimes \Phi^{\otimes n-1} + \Phi \otimes \Delta \otimes \Phi^{\otimes n-2} + \cdots + \Phi^{\otimes n-1} \otimes \Delta \in \mathcal{T}_\Phi\left(\mathcal{C}^{\otimes n}\right), \\
\{p, \delta\}^{\otimes n} &:= \left\{p^{\otimes n}, \delta^{(n)}\right\}, \\
\{\Phi, \Delta\}^{\otimes n} &:= \left\{\Phi^{\otimes n}, \Delta^{(n)}\right\}, \\
\{p_1, \delta_1\} \otimes \{p_2, \delta_2\} &:= \{p_1 \otimes p_2, \delta_1 \otimes p_2 + p_1 \otimes \delta_2\}.
\end{aligned}
$$

$\|\cdot\|_1$ denotes, for a (singed) measure, total variation, and for a function, L_1-norm. $\|\cdot\|_{\mathrm{cb}}$ denotes completely bounded norm: for a linear map Λ form signed measures (L^1-functions) to signed measures (L^1-functions),

$$
\|\Lambda\|_{\mathrm{cb}} = \max_{p:\ \text{probability distributions}} \|\Lambda \otimes \mathbf{I}(p)\|_1 .
$$

(Here note Λ may not be a Markov map, i.e., may not map a probability distirbution to another dsitribution.)

$g_p(\delta)$ and $G_\Phi(\Delta)$ denotes a metric, or square of a norm in $\mathcal{T}_p(\mathcal{P})$ and $\mathcal{T}_\Phi(\mathcal{C})$, respectively. In the present paper, they are not necessarily Riemanian. A probability distribution p is identified with the Markov map which sends all the input probability distributions to p, so that notations such as $G_p(\delta)$ makes sense. $J_p(\delta)$ denotes Fisher information,

$$
J_p(\delta) := \mathbb{E}\{L\}^2 = \int \{L(x)\}^2 p(x)\, \mathrm{d}\mu(x) = \int \frac{\{\delta(x)\}^2}{p(x)} \mathrm{d}\mu(x)
$$

Finally, $\Phi(\cdot|x) \in \mathcal{P}_{\mathrm{out}}$ is the distribution (, or its density) of the output when the input is x. Also, with $\Delta = c(\Phi_1 - \Phi_2)$,

$$
\Delta(\cdot|x) := c(\Phi_1(\cdot|x) - \Phi_2(\cdot|x)) \in \mathcal{T}_p(\mathcal{P}_{\mathrm{out}}).
$$

3 Probability distributions

Cencov had proven uniqueness (up to the constant multiple) of the monotone metric in the space of classical probability distributions defined over the finite set. In the proof, it is essential that the metric is Riemanian, i.e., induced from an inner product. As will be noted in Theorem 16, however, this assumption is not compatible with monotonicity in case of channels. Hence, we dispense with this assumption, and, instead, introduce new axioms which rules asymptotic behaviour of a metric.

3.1 Axioms for the metrics of probability distributions

(M0) $g_p(\delta) \geq g_{\Psi(p)}(\Psi(\delta))$.

(A0) $\lim_{n\to\infty}\frac{1}{n}\,g_{p^{\otimes n}}\left(\delta^{(n)}\right)=g_p\left(\delta\right)$.

(C0) If $\left\|q^n-p^{\otimes n}\right\|_1\to 0$ and $\frac{1}{\sqrt{n}}\left\|\delta'^n-\delta^{(n)}\right\|_1\to 0$ then

$$\varliminf_{n\to\infty}\frac{1}{n}\left(g_{q^n}\left(\delta'^n\right)-g_{p^{\otimes n}}\left(\delta^{(n)}\right)\right)\geq 0.$$

(N0) (Normalization) In case $\{p,\delta\}=\{\mathrm{N}\left(0,1\right),\delta\mathrm{N}\left(0,1\right)\}$ (a Gaussian shift family),

$$g_p\left(\delta\right)=1.$$

3.2 Simulation and asymptotic tangent smulation:definition

Simulation of $\{p_\theta\}$ is the pair $\{q_\theta,\Lambda\}$ with

$$p_\theta=\Lambda\left(q_\theta\right),\ \forall\theta\in\Theta,$$

and *tangent simulation* of the local data $\{p,\delta\}$ is the pair $\{q,\delta',\Lambda\}$ with

$$p=\Lambda\left(q\right),\ \ \delta=\Lambda\left(\delta'\right).$$

If in addition there is Λ' with

$$q=\Lambda'\left(p\right),\ \delta'=\Lambda'\left(\delta\right),$$

we say $\{p,\delta\}$ and $\{q,\delta'\}$ are *equivalent*, and express this relation by the notaton

$$\{p,\delta\}\equiv\{q,\delta'\}.$$

An *asymptotic tangent simulation* of $\left\{p^{\otimes n},\delta^{(n)}\right\}$ means a sequence $\{q^n,\delta'^n,\Lambda^n\}_{n=1}^{\infty}$ of triplet of a probability density q^n, an L^1-function δ'^n with $\int\delta'^n\mathrm{d}\mu=0$, and a Markov map Λ^n, such that

$$\lim_{n\to\infty}\left\|p^{\otimes n}-\Lambda^n\left(q^n\right)\right\|_1=0, \tag{1}$$

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\left\|\delta^{(n)}-\Lambda^n\left(\delta'^n\right)\right\|_1=0. \tag{2}$$

We call $\max\left\{\left\|p^{\otimes n}-\Lambda^n\left(q^n\right)\right\|_1,\left\|p^{\otimes n}-\Lambda^n\left(q^n\right)\right\|_1\right\}$ the *error* of the asymptotic tangent simulation. In all the cases treated int the present paper, the following stronger conditions are satisfied:

$$\left\|p^{\otimes n}-\Lambda^n\left(q^n\right)\right\|_1\leq\frac{1}{\sqrt{n}}C\left(\{p,\delta\}\right), \tag{3}$$

$$\frac{1}{\sqrt{n}}\left\|\delta^{(n)}-\Lambda^n\left(\delta'^n\right)\right\|_1\leq\frac{1}{n^{1/4}}C\left(\{p,\delta\}\right). \tag{4}$$

Below, $C\left(\{p,\delta\}\right)$ is sometimes denoted by C, as long as no confusion is likely to arise.

Proposition 1 *Let $L\left(x\right):=\delta\left(x\right)/p\left(x\right)$. Then,*

$$\{p,\delta\}\equiv\left\{\mathrm{p}_L\left(l\right),l\mathrm{p}_L\left(l\right)\right\}.$$

Proof. Observe

$$\int_{x:L(x)=l} p(x)\,\mathrm{d}\mu(x) = \mathrm{p}_L(l),$$
$$\int_{x:L(x)=l} \delta(x)\,\mathrm{d}\mu(x) = \int_{x:L(x)=l} L(x)\,p(x)\,\mathrm{d}\mu(x)$$
$$= l\int_{x:L(x)=l} p(x)\,\mathrm{d}\mu(x) = l\mathrm{p}_L(l),$$

where μ is either Lebesgue measure ($\Omega = \mathbb{R}^d$) or counting measure ($\Omega = \{1, \cdots, k\}$). Also,

$$\mathrm{p}_{X|L}(x|l)\,\mathrm{p}_L(l) = \begin{cases} p(x), & (l = L(x)) \\ 0, & \text{otherwise} \end{cases},$$
$$\mathrm{p}_{X|L}(x|l)\,\{l\mathrm{p}_L(l)\} = \begin{cases} L(x)\,p(x) = \delta(x), & (l = L(x)) \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, letting ν be a measure induced from μ via change of the variable $l = \delta(x)/p(x)$,

$$\int \mathrm{p}_{X|L}(x|l)\,\mathrm{p}_L(l)\,\mathrm{d}\nu(l) = p(x),$$
$$\int \mathrm{p}_{X|L}(x|l)\,\{l\mathrm{p}_L(l)\}\,\mathrm{d}\nu(l) = \delta(x).$$

■

Lemma 2 *Let $L'^n := \delta'^n/q^n$, and suppose that $q^n = \mathrm{p}_{L'^n}$. Let $\tilde{L}^n$ be a random variable defined over $\mathcal{B}\left(L^{(n)}\right)$, obeying the distribution*

$$\mathrm{p}_{\tilde{L}^n}(l^n) := \tilde{\Lambda}^n\left(\mathrm{p}_{L'^n}\right)(l^n)$$
$$:= \int P^n\left(l^n|l'^n\right)\mathrm{p}_{L'^n}\left(l'^n\right)\mathrm{d}l'^n.$$

Define Λ^n by

$$\Lambda^n(\mathrm{q})(x^n) := \int \tilde{\Lambda}^n(\mathrm{q})(l^n)\,\mathrm{p}_{X^n|L^{(n)}}(x^n|l^n)\,\mathrm{d}l^n.$$

For $\{q^n, \delta'^n, \Lambda^n\}_{n=1}^{\infty}$ to satisfy (3) and (4), it sufficeas that

$$\left\|\mathrm{p}_{L^{(n)}} - \mathrm{p}_{\tilde{L}^n}\right\|_1 \le \frac{C'}{\sqrt{n}}, \tag{5}$$

and

$$\max\left\{\mathbb{E}\left|\mathbb{E}\left[L'^n\,|\tilde{L}^n\right] - \tilde{L}^n\right|,\ \mathbb{E}(L)^2, \frac{1}{n}\mathbb{E}\left(\tilde{L}^n\right)^2\right\} \le a < \infty, \tag{6}$$

where

$$2C' + 3a \le C.$$

Proof. Since $\mathbb{P}_{X^n|L^{(n)}}\left(\mathcal{A}|l^{(n)}\right) \leq 1$, (5) implies

$$\left\|p^{\otimes n} - \Lambda^n\left(q^n\right)\right\|_1 = \sup_{\mathcal{A}:\text{measurable}} \left|\mathbb{E}\,\mathbb{P}_{X^n|L^{(n)}}\left(\mathcal{A}|L^{(n)}\right) - \mathbb{E}\,\mathbb{P}_{X^n|L^{(n)}}\left(\mathcal{A}|\tilde{L}^n\right)\right| \leq \frac{C'}{\sqrt{n}},$$

which is (3). By Chebychev's inequality,

$$\begin{aligned} \mathbb{E}\left\{\frac{1}{\sqrt{n}}\left|L^{(n)}\right| \;;\frac{1}{\sqrt{n}}\left|L^{(n)}\right| \geq n^{1/4}\right\} &\leq n^{-1/4}\cdot\frac{1}{n}\mathbb{E}\left(L^{(n)}\right)^2 \leq an^{-1/4}, \\ \mathbb{E}\left\{\frac{1}{\sqrt{n}}\left|\tilde{L}^n\right| \;;\frac{1}{\sqrt{n}}\left|\tilde{L}^n\right| \geq n^{1/4}\right\} &\leq an^{-1/4}. \end{aligned}$$

Also,

$$\begin{aligned} \tilde{\Lambda}^n\left(\delta'^n\right)\left(l^n\right) &= \int l'^n P^n\left(l^n|l'^n\right)\mathrm{p}_{L'^n}\left(l'^n\right)\mathrm{d}l'^n \\ &= \int l'^n \frac{P^n\left(l^n|l'^n\right)\mathrm{p}_{L'^n}\left(l'^n\right)}{\mathrm{p}_{\tilde{L}^n}\left(l^n\right)}\mathrm{d}l'^n\,\mathrm{p}_{\tilde{L}^n}\left(l^n\right) \\ &= \mathbb{E}\left[L^{'n}\,|\tilde{L}^n = l^n\right]\mathrm{p}_{\tilde{L}^n}\left(l^n\right). \end{aligned}$$

Therefore, by Proposition 1,

$$\begin{aligned} &\frac{1}{\sqrt{n}}\left\|\delta^{(n)} - \Lambda^n\left(\delta'^n\right)\right\|_1 \\ &\leq \frac{1}{\sqrt{n}}\int\left|l^n\mathrm{p}_{L^{(n)}}\left(l^n\right) - \tilde{\Lambda}^n\left(\delta'^n\right)\left(l^n\right)\right|\mathrm{d}l^n \\ &= \frac{1}{\sqrt{n}}\int\left|l^n\mathrm{p}_{L^{(n)}}\left(l^n\right) - \mathbb{E}\left[L^{'n}\,|\tilde{L}^n = l^n\right]\mathrm{p}_{\tilde{L}^n}\left(l^n\right)\right|\mathrm{d}l^n \\ &\leq \frac{1}{\sqrt{n}}\int\left|l^n\mathrm{p}_{L^{(n)}}\left(l^n\right) - l^n\mathrm{p}_{\tilde{L}^n}\left(l^n\right)\right|\mathrm{d}l^n \\ &+ \frac{1}{\sqrt{n}}\int\left|l^n\mathrm{p}_{\tilde{L}^n}\left(l^n\right) - \mathbb{E}\left[L^{'n}\,|\tilde{L}^n = l^n\right]\mathrm{p}_{\tilde{L}^n}\left(l^n\right)\right|\mathrm{d}l^n \\ &\leq \frac{1}{\sqrt{n}}\int\left|l^n\mathrm{p}_{L^{(n)}}\left(l^n\right) - l^n\mathrm{p}_{\tilde{L}^n}\left(l^n\right)\right|\mathrm{d}l^n + \frac{a}{n^{1/4}} \\ &\leq \mathbb{E}\left\{\frac{1}{\sqrt{n}}\mathbb{E}\left|L^{(n)} - \tilde{L}^n\right|;\frac{1}{\sqrt{n}}\left|L^{(n)}\right| \leq n^{1/4}, \frac{1}{\sqrt{n}}\left|\tilde{L}^n\right| \leq n^{1/4}\right\} + \frac{2a}{n^{1/4}} + \frac{a}{n^{1/4}} \\ &\leq 2n^{1/4}\left\|p^{\otimes n} - \Lambda^n\left(q^n\right)\right\|_1 + \frac{3a}{n^{1/4}} \\ &\leq \frac{3a+2C'}{n^{1/4}} \leq \frac{C}{n^{1/4}}. \end{aligned}$$

■

Proposition 3 *Suppose there is an asymptotic tangent simulation of* $\left\{p_{i+1}^n, \delta_{i+1}^n\right\}$ *by* $\left\{p_{i+1}^n, \delta_{i+1}^n\right\}$ *with the error* $f_i\left(n\right)$ *. Then, if* k *is a constant of* n*, there is an asymptotic tangent simulation of* $\left\{p_k^n, \delta_k^n\right\}$ *by* $\left\{p_1^n, \delta_1^n\right\}$ *with the error* $\sum_{i=1}^{k-1} f_i\left(n\right)$*.*

Proof. Obvious thus omitted. ■

3.3 Simulation of probabiltiy distribution family: a background from decision theory

Concept of simulation has been discussed in the field of statistical decision theory in relation with the notion of sufficiency [18]. Consider families $\mathcal{E}=\{p_\theta\}_{\theta\in\Theta}$, $\mathcal{F}=\{q_\theta\}_{\theta\in\Theta}$ of probability distributions, and a function $e:\theta\to e(\theta)>0$. Also, let $(D,\mathcal{D})$ be a decision space. Then $\mathcal{F}$ is said to be *e-deficient* relative to $\mathcal{E}$ if, for any loss function W_θ with $|W_\theta(d)|\leq 1$ and for any decision function $d:x\to d(x)\in D$, there is $d':y\to d'(y)\in D$ with

$$\int q_\theta(y)\,W_\theta(d'(y))\,\mathrm{d}\mu'\leq\int p_\theta(x)\,W_\theta(d(x))\,\mathrm{d}\mu+e_\theta. \tag{7}$$

0-defficiency is simply called deficiency. The celebrated *randomizing criteria*, a necessary and sufficient condition for e-defficiency is the existence of Λ with

$$\|p_\theta-\Lambda(q_\theta)\|\leq e_\theta.$$

Especially, 0-deficiency is equivalent to that $Y\sim q_\theta$ is a sufficient statistic of $\mathcal{E}=\{p_\theta\}_{\theta\in\Theta}$. Thus, e-defficiency is an approximate version of suffuciency.

This randomizing criteria motivates our emphasis on simulation. Its 'local' version

$$\sup_\theta\|p_\theta-\Lambda(q_\theta)\|=0$$
$$\left\|\frac{\partial p_\theta}{\partial\theta^i}-\Lambda\left(\frac{\partial q_\theta}{\partial\theta^i}\right)\right\|\leq e_i$$

is called *local e-deficiency at θ.*

3.4 Gaussian shift family

Proposition 4 *Suppose $\{p,\delta\}=\left\{\mathrm{N}\left(\theta,\sigma^2\right),\delta\mathrm{N}\left(\theta,\sigma^2\right)\right\}$. Suppose also (M0), and (N0) holds. Then we have*

$$g_p(\delta)=\frac{1}{\sigma^2}=J_p(\delta).$$

Proof. By an affine coordinate change of the data space $\Omega=\mathbb{R}$, $\{p,\delta\}=\{\mathrm{N}(\theta,\sigma),\delta\mathrm{N}(\theta,\sigma)\}$ is transformed to $\{q,\delta'\}=\left\{\mathrm{N}(0,1),\frac{1}{\sigma^2}\delta\mathrm{N}(0,1)\right\}$. Its inverse coordinate transform coordinate change of the data space Ω sends $\{q,\delta'\}$ to $\{p,\delta\}$. Therefore, by (M0) and (N0),

$$g_p(\delta)=g_{\mathrm{N}(0,1)}\left(\frac{1}{\sigma^2}\delta\mathrm{N}(0,1)\right)=\frac{1}{\sigma^2}g_{\mathrm{N}(0,1)}(\delta\mathrm{N}(0,1))=\frac{1}{\sigma^2}.$$

. ■

Remark 5 *Similarly, one can prove $\{\mathrm{N}(0,1),\delta\mathrm{N}(0,1)\}^{\otimes n}$, $\left\{\mathrm{N}\left(0,\frac{1}{n}\right),\delta\mathrm{N}\left(0,\frac{1}{n}\right)\right\}$, and $\{\mathrm{N}(0,1),\sqrt{n}\delta\mathrm{N}(0,1)\}$ are equivalent.*

3.5 On local asymptotic normality

Asymptotic tangent simulation by Gaussian shift is somewhat analogous to so-called *local asymptotic normality* (LAN, in short) [16]. Difference between them are as follows. First, asymptotic tangent simulation is concerned only with a particular point p, while LAN is concerned also with its neibourhood. On the other hand, (3) for asymptotic tangent simulation is norm convergence, and thus obviously stronger than convergence of $\frac{1}{\sqrt{n}}L^{(n)}$ to $\mathrm{N}(0,J)$ in law.

3.6 Zero bias transform

Let X be a real valued random variable with the distribution $\mathbb{P}_X$. Then,

$$W_X(x) := \frac{1}{\mathbb{V}(X)}\int_{-\infty}^{x}(\mathbb{E}X - y)\,\mathbb{P}_X(\mathrm{d}y)$$

satisfies $\int W_X(x)\,\mathrm{d}y = 1$, and thus defines a random variable X°. The map from X to X° is called called *zero-bias transform [3][14][13][7][8].* The following lemmas are proved in the literatures mentioned above.

Lemma 6 *Suppose $0 < \mathbb{V}(X) < \infty$ and $\mathbb{E}(X) = 0$. Suppose also $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous, differentiable, and $\mathbb{E}|f'(X^\circ)| < \infty$,*

$$\mathbb{E}(X f(X)) = \mathbb{V}(X)\,\mathbb{E}f'(X^\circ). \tag{8}$$

Lemma 7 *Let $S := \sum_{i=1}^{n} X_i$, where $X_1, \cdots, X_n$ are IID with $0 < \mathbb{V}(X_i) < \infty$ and $\mathbb{E}(X_i) = 0$. Then, denoting convolution by $*$,*

$$S^\circ = S - X_n + X_n^\circ,$$
$$W_S = W_X * (\mathrm{p}_X)^{*n-1}.$$

Lemma 8 *Let $S := a_1X_1 + a_2X_2$, where $a_1^2 + a_2^2 = 1$. If $W_{X_i}(x)/\mathrm{p}_{X_i}(x) < \infty$ and $\mathbb{E}\left(W_{X_i}(X)/\mathrm{p}_{X_i}(X) - 1\right)^2 < \infty$ $(i = 1, 2)$,*

$$\mathbb{E}\left(W_S(S)/\mathrm{p}_S(S) - 1\right)^2 \le a_1^4\mathbb{E}\left(W_{X_1}(X_1)/\mathrm{p}_{X_1}(X_1) - 1\right)^2 + a_2^4\mathbb{E}\left(W_{X_2}(X_2)/\mathrm{p}_{X_2}(X_2) - 1\right)^2$$

Lemma 9 *The random variable X° is supported on a subset of the convex hull of the support of X.*

3.7 Binary distributions

Consider a family of binary distributions $\{p_\theta\}$, where the data space is $\Omega = \{0, 1\}$. Letting $N_1(x^n)$ be the number of 1 in the sequence $x^n = x_1x_2\cdots x_n$,

$$\begin{aligned} L^{(n)} &= N_1(x^n)\{L(1) - L(0)\} + nL(0) \\ &= \alpha\{N_1(x^n) - np(1)\}, \end{aligned}$$

where $\alpha := L(1) - L(0)$.

We compose $\tilde{\Lambda}^n$ which satisfies (5) with $\left\{\mathrm{p}_{L^{(n)}},\, L^{(n)}\mathrm{p}_{L^{(n)}}\right\} \equiv \left\{p^{\otimes n}, \delta^{(n)}\right\} := \left\{p_\theta^{\otimes n}, \delta_\theta^{(n)}\right\}$ and

$$\begin{aligned}\{q^n, \delta'^n\} &:= \{\mathrm{N}(0, nJ_p(\delta)), nJ_p(\delta)\,\delta\mathrm{N}(0, nJ_p(\delta))\} \\ &\equiv \left\{\mathrm{N}(0,1), \sqrt{nJ_p(\delta)}\delta\mathrm{N}(0,1)\right\} \equiv \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes nJ_p(\delta)},\end{aligned}$$

by letting $\tilde{L}^n$ be the element of the set

$$\{\alpha(n_1 - p(1)n)\ ; n_1 \in \mathbb{N}, n_1 \le n\}$$

closest to $L'^n \sim \mathrm{N}(0, nJ_p(\delta))$.

One can easily verify

$$\begin{aligned}\mathbb{E}\left|\mathbb{E}\left[L'^n\,|\tilde{L}^n\right] - \tilde{L}^n\right| &\le |\alpha_\theta|, \\ \mathbb{E}(L)^2 &= J_p(\delta) < \infty, \\ \frac{1}{n}\mathbb{E}\left(\tilde{L}^n\right)^2 &\le \frac{1}{n}\mathbb{E}(L'^n)^2 + \frac{1}{n}\mathbb{E}\left|L'^n - \tilde{L}^n\right|^2 \\ &\le J_p(\delta) + \frac{1}{n}(\alpha)^2.\end{aligned}$$

(5), or

$$\left\|p^{\otimes n} - \Lambda(q^n)\right\|_1 \le \left\|\mathrm{p}_{L^{(n)}} - \mathrm{p}_{\tilde{L}^n}\right\|_1 \le \frac{1}{\sqrt{n}}\frac{4}{\sqrt{J_p(\delta)}},$$

is the direct consequence of Theorem 10 below. Hence, by Lemma 2, the error of this tangent simulation is $\frac{A}{n^{1/4}}$ with

$$A = \frac{8}{\sqrt{J_p(\delta)}} + 3J_p(\delta) + 3(\alpha)^2 + 3|\alpha|,$$

which is continuous function of $\delta(0)$ and $p(0)$ is bounded on any compact region.

Theorem 10 *Let $X_1, X_2, \cdots, X_n$ be the IID random variables taking values in $\{0,1\}$, with $\Pr\{X_1 = 1\} = \eta$. Denote its variance by σ^2, and define*

$$Y_i := \frac{1}{\sqrt{n}\sigma}(X_i - \eta), \quad S_n := \sum_{i=1}^n Y_i.$$

Suppose

$$\mathcal{A} = \bigcup_z \mathcal{A}_z^n,$$

where

$$\mathcal{A}_z^n := \left[z - \frac{1}{2\sqrt{n}\sigma}, z + \frac{1}{2\sqrt{n}\sigma}\right]$$

and z runs over a subset of $(\mathbb{Z} - n\eta)/\sqrt{n}\sigma$. Then,

$$|\Pr\{S_n \in \mathcal{A}\} - \Pr\{\mathrm{N}(0,1) \in \mathcal{A}\}| \le \frac{1}{\sqrt{n}\sigma}.$$

Proof. Letting

$$\psi_{\mathcal{A}}(x) := e^{\frac{x^2}{2}} \int_{-\infty}^{x} \left(\chi_{\mathcal{A}}(t) - \Pr\{\mathrm{N}(0,1) \in \mathcal{A}\}\right) e^{-\frac{t^2}{2}} \mathrm{d}t, \tag{9}$$

we have

$$\begin{aligned}
&\left|\Pr\{S_n \in \mathcal{A}\} - \Pr\{\mathrm{N}(0,1) \in \mathcal{A}\}\right| \underset{(i)}{=} \left|\mathbb{E}\left(\frac{\mathrm{d}}{\mathrm{d}x}\psi_{\mathcal{A}}(S_n) - S_n \psi_{\mathcal{A}}(S_n)\right)\right| \\
&\underset{(ii)}{=} \left|\mathbb{E}\left(\frac{\mathrm{d}}{\mathrm{d}x}\psi_{\mathcal{A}}(S_n) - \frac{\mathrm{d}}{\mathrm{d}x}\psi_{\mathcal{A}}(S_n^\circ)\right)\right| \\
&\underset{(iii)}{=} \left|\mathbb{E}\left(\chi_{\mathcal{A}}(S_n) - \chi_{\mathcal{A}}(S_n^\circ)\right) + \mathbb{E}\left(S_n \psi_{\mathcal{A}}(S_n) - S_n^\circ \psi_{\mathcal{A}}(S_n^\circ)\right)\right| \\
&\underset{(iv)}{=} 2\left|\mathbb{E}\left(\chi_{\mathcal{A}}(S_n) - \chi_{\mathcal{A}}(S_n^\circ)\right)\right|, \\
&\underset{(v)}{=} 2\left|\mathbb{E}\left(\chi_{\mathcal{A}}(S_n) - \chi_{\mathcal{A}}(S_n - Y_n + Y_n^\circ)\right)\right|.
\end{aligned} \tag{10}$$

where (i) and (iii) are due to the definition (9), (ii) is due to (8), (iv) is due to

$$\psi_{\mathcal{A}}(x) \leq \frac{1}{|x|},$$

and (v) is due to Lemma 7. By definition, one can verify that $X_i^\circ \sim W_{X_i}$ is uniform distirbution over $[0,1]$. Also,

$$\begin{aligned}
&\left|\mathbb{E}\left(\chi_{\mathcal{A}}(S_n) - \chi_{\mathcal{A}}(S_n - Y_n + Y_n^\circ)\right)\right| \\
&\leq \sum_{k=0}^{n} \left|\Pr\left\{\sum_{i=1}^{n-1} X_i + X_n = k\right\} - \Pr\left\{\sum_{i=1}^{n-1} X_i + X_n^\circ \in \left[k - \frac{1}{2}, k + \frac{1}{2}\right]\right\}\right| \\
&= \sum_{k=0}^{n} \left|\Pr\left\{\sum_{i=1}^{n-1} X_i = k\right\}\left\{(1-\eta) - \frac{1}{2}\right\} + \Pr\left\{\sum_{i=1}^{n-1} X_i = k-1\right\}\left(\eta - \frac{1}{2}\right)\right| \\
&= \sum_{k=0}^{n} \Pr\left\{\sum_{i=1}^{n} X_i = k\right\}\left|\frac{1}{1-\eta}\frac{n-k}{n}\left\{(1-\eta) - \frac{1}{2}\right\} + \frac{1}{\eta}\frac{k}{n}\left(\eta - \frac{1}{2}\right)\right| \\
&= \frac{\left|\eta - \frac{1}{2}\right|}{\eta(1-\eta)} \sum_{k=0}^{n} \Pr\left\{\sum_{i=1}^{n} X_i = k\right\}\left|\frac{k}{n} - \eta\right| \\
&\leq \frac{\left|\eta - \frac{1}{2}\right|}{\eta(1-\eta)} \sqrt{\sum_{k=0}^{n} \Pr\left\{\sum_{i=1}^{n} X_i = k\right\}\left(\frac{k}{n} - \eta\right)^2} = \frac{\left|\eta - \frac{1}{2}\right|}{\{\eta(1-\eta)\}^{1/2}} \frac{1}{\sqrt{n}},
\end{aligned}$$

which leads to the assertion. ■

3.8 Distributions over the finite set

Theorem 11 *Suppose p is a probability distribution and δ is a signed measure over a set Ω with $|\Omega| = k$ $(k < \infty)$. Let $J := J_p(\delta)$, $\varepsilon > 0$ and*

$$\{q^n, \delta'^n\} := \left\{\mathrm{N}(0,1), \sqrt{n(J+\varepsilon)}\delta\mathrm{N}(0,1)\right\} \equiv \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n(J+\varepsilon)}.$$

Then, we can compose Λ^n with the error $\frac{A}{n^{1/4}}$, where A is a continuous function of $\{p(x), \delta(x); x = 1, \cdots, k-1\}$ and is bounded on any compact region.

Proof. Since binary distributions can be simulated by Gaussian shift as in Subsection 3.7, due to Propositin 3, we only have to compose asymptotic tangent simulation $\{p, \delta\}^{\otimes n}$ by binary distributions. For that, we first asymptotically simulate $\{p, \delta\}^{\otimes n}$ by $\{p_a, \delta_a\}^{\otimes n} \otimes \{p_A, \delta_A\}^{\otimes n_a}$, where $\{p_a, \delta_a\}$ and $\{p_A, \delta_A\}$ is defined over the binary set Ω_a and and the set Ω_A with $(k-1)$-elements, respectively.Then, by virtue of Propositin 3, inductive argument leads to asymptotic tangent simulation by binary distributions.

Let

$$
\begin{aligned}
p_a(0) &:= p(k), \ p_a(1) = \sum_{x=1}^{k-1} p(x), \\
\delta_a(0) &:= \delta(k), \ \delta_a(1) = \sum_{x=1}^{k-1} \delta(x), \ L_a := \frac{\delta_a}{p_a} \\
p_A(x) &:= \frac{p(x)}{p_a(1)}, \ (x = 1, \cdots, k-1), \\
\delta_A(x) &:= \frac{\delta(x)}{p_a(1)} - \frac{\delta_a(1) p_A(x)}{p_a(1)}, \ (x = 1, \cdots, k-1), \\
L_A &:= \frac{\delta_A(x)}{p_A(x)}, \quad (x = 1, \cdots, k-1), \\
n_a &:= n(p_a(1) + \varepsilon) \quad (\varepsilon > 0).
\end{aligned}
$$

Also, let $x_a^n = x_{a1} x_{a2} \cdots x_{an} \in \Omega_a^{\otimes n}$, $x_A^n = x_{A1} x_{A2} \cdots x_{An} \in \Omega_A^{\otimes n}$, $X_{ai} \sim p_a$, $X_{Ai} \sim p_A$, $X_a^n \sim p_a^{\otimes n}$, and $X_A^n \sim p_A^{\otimes n}$. Denote by $N_1(x_a^n)$ the number of 1 in the sequence $x_a^n = x_{a1} x_{a2} \cdots x_{an}$. Also, we identify the pair (x_a, x_A) with x, by the correspondence

$$
x \equiv \begin{cases} (1, x) & (x = 1, \cdots, k-1), \\ (0, \#) & x = k, \end{cases}
$$

where # stands for empty string. To define asymptotic tangent simulation, one define function $F : \Omega^{\otimes n} \to \Omega^{\otimes n}$ such that

$$
F(x^n) = \begin{cases} x^n & (N_1(x_a^n) \le n_a), \\ k^n & (N_1(x_a^n) > n_a). \end{cases}
$$

Using F, we define

$$
\Lambda^n(r^n)(x^n) := \sum_{y^n \in F^{-1}(x^n)} r^n(y^n),
$$

$$
\tilde{p}^n := \Lambda^n\left(p^{\otimes n}\right), \ \tilde{\delta}^n := \Lambda^n\left(\delta^{(n)}\right).
$$

Then,

$$\left\|\tilde{p}^{n}-p^{\otimes n}\right\|_{1}=\sum_{x^{n}}\left|\sum_{y^{n}\in F^{-1}(x^{n})}p^{\otimes n}\left(y^{n}\right)-p^{\otimes n}\left(x^{n}\right)\right|$$
$$=2\sum_{x^{n}:N_{1}(x_{a}^{n})>n_{a}}p^{\otimes n}\left(x^{n}\right)$$
$$\leq 2\exp\left\{-nC_{a,\varepsilon}\right\},$$

where

$$C_{a,\varepsilon}:=p_{a}\left(0\right)\ln\frac{p_{a}\left(0\right)}{p_{a}\left(0\right)-\varepsilon}+p_{a}\left(1\right)\ln\frac{p_{a}\left(1\right)}{p_{a}\left(1\right)+\varepsilon},$$

and

$$\begin{aligned}
\frac{1}{\sqrt{n}}\left\|\tilde{\delta}^{n}-\delta^{(n)}\right\|_{1} &= \frac{1}{\sqrt{n}}\left|\sum_{N_{1}(x_{a}^{n})>n_{a}}\delta^{(n)}\left(x^{n}\right)\right|+\frac{1}{\sqrt{n}}\sum_{N_{1}(x_{a}^{n})>n_{a}}\left|\delta^{(n)}\left(x^{n}\right)\right| \\
&\leq \frac{2}{\sqrt{n}}\sum_{N_{1}(x_{a}^{n})>n_{a}}\left|\delta^{(n)}\left(x^{n}\right)\right| \\
&= \frac{2}{\sqrt{n}}\sum_{x_{a}^{n}:N_{1}(x_{a}^{n})>n_{a}}\sum_{x_{A}^{N_{A}(x_{a}^{n})}}\left|L_{a}^{(n)}\left(x_{a}^{n}\right)+L_{A}^{(N_{1}(x_{a}^{n}))}\left(x_{A}^{N_{1}(x_{a}^{n})}\right)\right|p_{a}^{\otimes n}\left(x_{a}^{n}\right)p_{A}^{\otimes N_{A}(x_{a}^{n})}\left(x_{A}^{N_{A}(x_{a}^{n})}\right) \\
&\leq \frac{2}{\sqrt{n}}\sum_{x_{a}^{n}:N_{1}(x_{a}^{n})>n_{a}}\left[n\max\left\{\left|L_{a}\left(0\right)\right|,\left|L_{a}\left(1\right)\right|\right\}+n\left|\max_{1\leq x_{A1}\leq k-1}L_{A}\left(x_{A1}\right)\right|\right]p_{a}^{\otimes n}\left(x_{a}^{n}\right) \\
&\leq 2\sqrt{n}\left[\max\left\{\left|L_{a}\left(0\right)\right|,\left|L_{a}\left(1\right)\right|\right\}+\left|\max_{1\leq x\leq k-1}L_{A}\left(x\right)\right|\right]\exp\left\{-nC_{a,\varepsilon}\right\}.
\end{aligned}$$

Also,

$$\begin{aligned}
J_{p}\left(\delta\right) &= \frac{\left\{\delta\left(k\right)\right\}^{2}}{p\left(k\right)}+\sum_{x=1}^{k-1}\frac{\left\{\delta\left(x\right)\right\}^{2}}{p\left(x\right)} \\
&= \frac{\left\{\delta_{a}\left(0\right)\right\}^{2}}{p_{a}\left(0\right)}+\sum_{x=1}^{k-1}\frac{1}{p_{a}\left(1\right)p_{A}\left(x\right)}\left\{p_{a}\left(1\right)\delta_{A}\left(x\right)+\delta_{a}\left(1\right)p_{A}\left(x\right)\right\}^{2} \\
&= \frac{\left\{\delta_{a}\left(0\right)\right\}^{2}}{p_{a}\left(0\right)}+\frac{\left\{\delta_{a}\left(1\right)\right\}^{2}}{p_{a}\left(1\right)}+p_{a}\left(1\right)\sum_{x=1}^{k-1}\frac{\left\{\delta_{A}\left(x\right)\right\}^{2}}{p_{A}\left(x\right)}+\delta_{a}\left(1\right)\sum_{x=1}^{k-1}\delta_{A}\left(x\right) \\
&= J_{a}\left(\delta_{a}\right)+p_{a}\left(1\right)J_{A}\left(\delta_{A}\right).
\end{aligned}$$

■

Analogously, one can compose an asymptotic tangent simulation of $\left\{p_{A},\delta_{A}\right\}^{\otimes n_{a}}$ by $\left\{p_{b},\delta_{b}\right\}^{\otimes n_{a}}\otimes\left\{p_{B},\delta_{B}\right\}^{\otimes n_{b}}$, where $\left\{p_{b},\delta_{b}\right\}$ and $\left\{p_{B},\delta_{B}\right\}$ are defined over the binary set Ω_{b} and and the set Ω_{B} with $(k-2)$-elements ,

respectively, where

$$
\begin{aligned}
p_b(0) &:= p_A(k-1) \\
p_b(1) &:= \sum_{x=1}^{k-2} p_A(x) \\
n_b &:= n_a\left(p_b(1)+\varepsilon\right), \\
J_{p_A}(\delta_A) &= J_{p_b}(\delta_b) + p_b(1)\, J_{p_B}(\delta_B).
\end{aligned}
$$

Repeating this proscess recursively, by Proposition 3, one can asymptotically simulate $\{p,\delta\}^{\otimes n}$ by

$$
\{p_a,\delta_a\}^{\otimes n} \otimes \{p_b,\delta_b\}^{\otimes n_a} \otimes \cdots \otimes \{p_z,\delta_z\}^{\otimes n_y}, \tag{11}
$$

($\{p_Z,\delta_Z\}$ is defined over $\{1\}$, thus is trivia)l with the error

$$
2\sqrt{n}\left[\sum_{i=a}^{z} \max\left\{|L_i(0)|, |L_i(1)|\right\} + \sum_{j=A}^{Y}\left|\max_{1\le x\le k-1} L_j(x)\right| + 1\right] \sum_{i=a}^{z} \exp\left\{-nC_{i,\varepsilon}\right\},
$$

which is upperbouded by $\frac{B}{n^{1/4}}$, where B is a continuous function of $\{p(x),\delta(x); x=1,\cdots,k-1\}$. Due to Subsection 3.7 , (11) can be simulated by

$$
\begin{aligned}
&\{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes nJ_a(\delta_a)} \otimes \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n_a J_b(\delta_b)} \otimes \\
&\cdots \otimes \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n_y J_z(\delta_z)} \\
&\equiv \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n(J_p(\delta)+f(\varepsilon))},
\end{aligned}
$$

where $\lim_{\varepsilon\to o} f(\varepsilon) = 0$, with the error $\frac{B'}{n^{1/4}}$, where B' is a continuous function of $\{p(x),\delta(x); x=1,\cdots,k-1\}$. Here, '$\equiv$' is due to

$$
\begin{aligned}
&nJ_a(\delta_a) + n_a J_b(\delta_b) + n_b J_c(\delta_c) + \cdots + n_y J_z(\delta_z) \\
&= nJ_a(\delta_a) + n\left(p_a(1)+\varepsilon\right) J_b(\delta_b) + n\left(p_a(1)+\varepsilon\right)\left(p_b(1)+\varepsilon\right) J_c(\delta_c) \\
&+ \cdots + n\prod_{i=a}^{y}\left(p_i(1)+\varepsilon\right) J_z(\delta_z) \\
&= n\left(J_p(\delta) + f(\varepsilon)\right)
\end{aligned}
$$

where the last identiy is due to

$$
J_p(\delta) = J_a(\delta_a) + p_a(1)\, J_b(\delta_b) + p_a(1)\, p_b(1)\, J_c(\delta_c) + \cdots + \prod_{i=a}^{y} p_i(1)\, J_z(\delta_z).
$$

Therefore, due to Proposition 3, we obtain an asymptotic tangent simulation of $\{p,\delta\}^{\otimes n}$ by $\{q^n,\delta'^n\}$ with the error $\frac{B+B'}{n^{1/4}}$, and the assertion is proved.

3.9 A continuous random variable with smooth density

Theorem 12 *Let $\Omega=\mathbb{R}$. Suppose $L(x)=\delta(x)/p(x)$ exists and is a continuous function of x. Let $J:=J_p(\delta)$, $\{q^n,\delta'^n\}:=\left\{\mathrm{N}(0,1),\sqrt{nJ}\delta\mathrm{N}(0,1)\right\}=\{\mathrm{N}(0,1),\delta\mathrm{N}(0,1)\}^{\otimes nJ}$, and define $\tilde{L}^n=\Lambda^{n*}(L'^n):=L'^n$. Suppose*

$$\mathbb{E}\left(\frac{\int_{-\infty}^{L}(-t)\ \mathrm{p}_L(t)\,\mathrm{d}t}{J\,\mathrm{p}_L(L)}\right)^2<\infty, \tag{12}$$

holds. Then, $\{q^n,\delta'^n,\Lambda^n\}$ satisfies (5) and (6). Thus, by Lemma 2, it satisfies (3) and (4).

Proof. The assertion is essentially the same as Theorem 2.3 of [14]. For the sake of completeness, however, the whole argument is described below. Let $S_n:=\frac{1}{\sqrt{nJ}}L^{(n)}$. In the same way as the proof of Theorem 10, we have

$$\begin{aligned}
&|\Pr\{S_n\in\mathcal{A}\}-\Pr\{\mathrm{N}(0,1)\in\mathcal{A}\}|\\
&=|\mathbb{E}(\chi_{\mathcal{A}}(S_n)-\chi_{\mathcal{A}}(S_n^\circ))+\mathbb{E}(S_n\psi_{\mathcal{A}}(S_n)-S_n^\circ\psi_{\mathcal{A}}(S_n^\circ))|\\
&\leq 2\left\|\mathrm{p}_{S_n}-\mathrm{p}_{S_n^\circ}\right\|_1=2\mathbb{E}\left|1-\frac{W_{S_n}(S_n)}{\mathrm{p}_{S_n}(S_n)}\right|\leq 2\sqrt{\mathbb{E}\left(\frac{W_{S_n}(S_n)}{\mathrm{p}_{S_n}(S_n)}-1\right)^2}.
\end{aligned}$$

where $\psi_{\mathcal{A}}$ is defined by (9), and thus $|x\psi_{\mathcal{A}}(x)|\leq 1$. Hence, it boils down to the evaluation of $\mathbb{E}\left(\frac{W_{S_n}(S_n)}{\mathrm{p}_{S_n}(S_n)}-1\right)^2$, which, due to Lemma 8, is not larger than

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\mathbb{E}\left(\frac{W_{\frac{1}{\sqrt{J}}L}\left(\frac{1}{\sqrt{J}}L\right)}{\mathrm{p}_{\frac{1}{\sqrt{J}}L}\left(\frac{1}{\sqrt{J}}L\right)}-1\right)^2=\frac{1}{\sqrt{n}}\mathbb{E}\left(\frac{\int_{-\infty}^{\frac{1}{\sqrt{J}}L}(-t)\ \mathrm{p}_{\frac{1}{\sqrt{J}}L}(t)\,\mathrm{d}t}{\mathrm{p}_{\frac{1}{\sqrt{J}}L}\left(\frac{1}{\sqrt{J}}L\right)}-1\right)^2\\
&=\frac{1}{\sqrt{n}}\mathbb{E}\left(\frac{\int_{-\infty}^{L}\left(-\frac{1}{\sqrt{J}}t\right)\ \mathrm{p}_L(t)\,\mathrm{d}t}{\sqrt{J}\mathrm{p}_L(l)}-1\right)^2=\frac{1}{\sqrt{n}}\mathbb{E}\left(\frac{\int_{-\infty}^{L}(-t)\ \mathrm{p}_L(t)\,\mathrm{d}t}{J\,\mathrm{p}_L(l)}-1\right)^2\\
&=\frac{1}{\sqrt{n}}\left\{\mathbb{E}\left(\frac{\int_{-\infty}^{L}(-t)\ \mathrm{p}_L(t)\,\mathrm{d}t}{J\,\mathrm{p}_L(l)}\right)^2-1\right\}.
\end{aligned}$$

Hence, we have (5). Also, it is easy to verify

$$\begin{aligned}
\mathbb{E}\left|\mathbb{E}\left[L'^n\,|\tilde{L}^n\right]-\tilde{L}^n\right|&=0,\\
\mathbb{E}(L)^2=\frac{1}{n}\mathbb{E}\left(\tilde{L}^n\right)^2&=J_p(\delta).
\end{aligned}$$

■

A trivial sufficient condition for (12) is that the support of p_L is bounded. Also, suppose

$$\begin{aligned}
\frac{a_1}{t^{\alpha_1}}&\leq\mathrm{p}_L(t)\leq\frac{b_1}{t^{\alpha_1}},\ (t\leq\exists t_1)\\
\frac{a_2}{t^{\alpha_2}}&\leq\mathrm{p}_L(t)\leq\frac{b_2}{t^{\alpha_2}},(t\geq\exists t_2)
\end{aligned}$$

hold for some real constant a_i, b_i, α_i ($\iota = 1{,}2$). Then, if $y < t_1$,

$$\frac{1}{\mathrm{p}_L(y)} \int_{-\infty}^{y} (-t)\, \mathrm{p}_L(t)\, \mathrm{d}t \le \frac{1}{\alpha_1 - 2} \frac{b_1}{a_1} y^2,$$

and, due to $\int_{-\infty}^{\infty} (-t)\, \mathrm{p}_L(t)\, \mathrm{d}t = 0$, if $y > t_2$,

$$\frac{1}{\mathrm{p}_L(y)} \int_{-\infty}^{y} (-t)\, \mathrm{p}_L(t)\, \mathrm{d}t = \frac{1}{\mathrm{p}_L(y)} \int_{y}^{\infty} t\, \mathrm{p}_L(t)\, \mathrm{d}t \le \frac{1}{\alpha_2 - 2} \frac{b_2}{a_2} y^2.$$

Hence, if

$$\min\{\alpha_1, \alpha_2\} \ge 4,$$

we have (12).

The following conditions are also sufficient:

$$\begin{aligned} a_1 e^{-|t|^{\alpha_1}} &\le \mathrm{p}_L(t) \le b_1 e^{-|t|^{\alpha_1}}, \ (t \le \exists t_1), \\ a_2 e^{-|t|^{\alpha_2}} &\le \mathrm{p}_L(t) \le b_2 e^{-|t|^{\alpha_2}}, (t \ge \exists t_2) \end{aligned} \tag{13}$$

for some real constants a_i, b_i, and α_i ($i = 1{,}2$) , with

$$\min\{\alpha_1, \alpha_2\} \ge 2.$$

Then, if $y < t_1$ and $y \le -1$,

$$\begin{aligned} \frac{1}{\mathrm{p}_L(y)} \int_{-\infty}^{y} (-t)\, \mathrm{p}_L(t)\, \mathrm{d}t &\le \frac{b_1}{a_1 e^{-|y|^{\alpha_1}}} \int_{-\infty}^{y} (-t)\, e^{-|t|^{\alpha_1}} \mathrm{d}t \\ &\le \frac{b_1}{a_1 e^{-|y|^{\alpha_1}}} \int_{-\infty}^{y} (-t)^{\alpha_1 - 1} e^{-|t|^{\alpha_1}} \mathrm{d}t \\ &= \frac{b_1}{a_1 e^{-|y|^{\alpha_1}}} \frac{e^{-|y|^{\alpha_1}}}{\alpha_1 - 1} = \frac{b_1}{a_1} \frac{1}{\alpha_1 - 1}. \end{aligned}$$

Hence, if $|y|$ is large enough enough, we have

$$\frac{1}{\mathrm{p}_L(y)} \int_{-\infty}^{y} (-t)\, \mathrm{p}_L(t)\, \mathrm{d}t < const.$$

The same is true for $y > t_2$-case, and thus (13) is another sufficient condition for (12).

3.10 Simulation of Gaussian shift by an arbitrary IID sequence

Suppose $\{q^n, \delta'^n\} = \left\{q^{\otimes n}, \delta'^{(n)}\right\}$, where $J_q(\delta') = J$, is given. Suppose also that $L' := \frac{\delta'}{q}$ has density with respect to Lebesgue measure, and satisfies (12). Then, by Theorem 12, we can compose asymptotic tangent simulation of

$$\left\{p^{\otimes n}, \delta^{(n)}\right\} := \{\mathrm{N}(0,1), \delta \mathrm{N}(0,1)\}^{\otimes nJ} \equiv \left\{\mathrm{N}(0,1), \sqrt{nJ}\delta \mathrm{N}(0,1)\right\}$$

with $\{q^n, \delta'^n\} = \left\{q^{\otimes n}, \delta'^{(n)}\right\}$.

Meanwhile, instead, suppose $|L'| \leq const.$ with probability 1. Then, by a given Let $X^i \sim q$, and $Y^i \sim \mathrm{N}(0,1)$,

$$\frac{1}{\sqrt{n}}\tilde{L}^n = \frac{1}{\sqrt{n}}(\Lambda^n)^*\left(L'^{(n)}\right) := \frac{1}{\sqrt{n(J+\varepsilon^2)}}\sum_{i=1}^{n}\left(L'\left(X^i\right)+\varepsilon Y^i\right).$$

Then $L'\left(X^i\right)+\varepsilon Y^i$ has density with respect to Lebesgue measure, and satisfies (13). Since Fisher information of $\mathrm{p}_{\tilde{L}^n}$ equals

$$\frac{n}{J^{-1}+\varepsilon^2} = \frac{nJ}{1+\varepsilon^2 J} = n\left(J - f(\varepsilon)\right) \quad (\lim_{\varepsilon\to 0} f(\varepsilon) = 0),$$

by Theorem 12, one can compose an asymptotic tangent symulation of

$$\left\{\mathrm{N}(0,1), \sqrt{n(1-f(\varepsilon))J}\,\delta\mathrm{N}(0,1)\right\} \equiv \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n(1-f(\varepsilon))}$$

by $\{\mathrm{p}_{\tilde{L}^n}(l), l\mathrm{p}_{\tilde{L}^n}(l)\}$. Since $\{\Lambda^n, \mathrm{p}_{L'^{(n)}}(l), l\mathrm{p}_{L'^{(n)}}(l)\}$ is an asymptotic tangent symulation of $\{\mathrm{p}_{\tilde{L}^n}(l), l\mathrm{p}_{\tilde{L}^n}(l)\}$, by Proposition 1 and Proposition 3, one can compose an asymptotic tangent simulation of $\{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n(1-f(\varepsilon))}$ with $\{q^n, \delta'^n\} = \left\{q^{\otimes n}, \delta'^{(n)}\right\}$.

3.11 Uniqueness theorem

Theorem 13 *Suppose g satisfies (M0), (A0), (C0), and (N0). Suppose also either (a): $\{p, \delta\}$ is defined over a finite set, or (b): the probability density p_L of L with respect to Lebesgue measure exists and satisfies (12). Then, if $\mathbb{E}_p(L)^4 < \infty$, $g_p(\delta)$ equals $J = J_p(\delta)$.*

Proof. Let $\{q^n, \delta'^n\} := \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n(J+\varepsilon)} = \left\{\mathrm{N}(0,1), \sqrt{n(J+\varepsilon)}\delta\mathrm{N}(0,1)\right\}$ $(\varepsilon > 0)$. Then by Proposition 4,

$$g_{q^n}(\delta'^n) = n(J+\varepsilon).$$

Due to Theorem 11 and Theorem 12, there is Λ^n with (3) and (4). Therefore, by (C0) and (M0),

$$\begin{aligned}
0 &\leq \varliminf_{n\to\infty}\frac{1}{n}\left(g_{\Lambda(q^n)}(\Lambda(\delta'^n)) - g_{p^{\otimes n}}\left(\delta^{(n)}\right)\right) \\
&\leq \varliminf_{n\to\infty}\frac{1}{n}\left(g_{q^n}(\delta'^n) - g_{p^{\otimes n}}\left(\delta^{(n)}\right)\right) \\
&= J_p(\delta) + \varepsilon - \varlimsup_{n\to\infty}\frac{1}{n}g_{p^{\otimes n}}\left(\delta^{(n)}\right).
\end{aligned}$$

Similarly, by the argument in Subsection 3.10, we have,

$$\begin{aligned}
0 &\leq \varliminf_{n\to\infty}\frac{1}{n}\left(g_{\Lambda(p^{\otimes n})}\left(\Lambda\left(\delta^{(n)}\right)\right) - g_{\mathrm{N}(0,1)}\left(, \sqrt{n(J-\varepsilon)}\delta\mathrm{N}(0,1)\right)\right) \\
&\leq \varliminf_{n\to\infty}\frac{1}{n}\left(g_{p^{\otimes n}}\left(\delta^{(n)}\right) - (J_p(\delta) - \varepsilon)\right)
\end{aligned}$$

Therefore,

$$J_p(\delta) - \varepsilon \leq \varliminf_{n\to\infty}\frac{1}{n}g_{p^{\otimes n}}\left(\delta^{(n)}\right) \leq \varlimsup_{n\to\infty}\frac{1}{n}g_{p^{\otimes n}}\left(\delta^{(n)}\right) \leq J_p(\delta) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have

$$\lim_{n\to\infty} \frac{1}{n} g_{p^{\otimes n}} \left(\delta^{(n)}\right) = J_p\left(\delta\right),$$

which, combined with (A0) implies

$$g_p\left(\delta\right) = J_p\left(\delta\right).$$

We have to check $g_p(\delta) = J_p(\delta)$ satisfies (M0), (A0), (C0), and (N0). (A0) and (N0) are checked by easy computation. (M0) is well-known. Hence, (C0) is shown in the sequel. We use the following characterization of Fisher information (see Chap. 9 of [2]):

$$J_p\left(\delta\right) = \max_T \frac{\left|\mathbb{E}_p LT\right|^2}{\mathbb{E}_p T^2},$$

where the maximum is achieved by $T = J^{-1} \cdot L$, with $J = J_p(\delta)$. Define $T^n := (nJ)^{-1} \cdot L^{(n)}$, and

$$T_a\left(x\right) := \begin{cases} T\left(x\right), & \left(\left|T\left(x\right)\right| \le a\right), \\ 0, & \left(\left|T\left(x\right)\right| > a\right), \end{cases}$$

$$T_a^n\left(x\right) := \begin{cases} T^n\left(x^n\right), & \left(T^n\left(x^n\right) \le a\right), \\ 0, & \left(T^n\left(x^n\right) > a\right). \end{cases}$$

Observe

$$\frac{1}{n} J_{q^n}\left(\delta'^n\right) \ge \frac{\left|\frac{1}{\sqrt{n}} \mathbb{E}_{q^n} L'^n T_a^n\right|^2}{\mathbb{E}_{q^n}\left(T_a^n\right)^2} = \frac{\left|\frac{1}{\sqrt{n}} \mathbb{E}_{p^{\otimes n}} L^{(n)} T_a^n\right|^2}{\mathbb{E}_{p^{\otimes n}}\left(T_a^n\right)^2} + o\left(1\right),$$

where the last identity is due to $\left\|q^n - p^{\otimes n}\right\|_1 \to 0$ and $\frac{1}{\sqrt{n}}\left\|\delta'^n - \delta^{(n)}\right\|_1 \to 0$. Observe also

$$\begin{aligned} \left|\mathbb{E}_{p^{\otimes n}}\left(T_a^n\right)^2 - \mathbb{E}_{p^{\otimes n}}\left(T^n\right)^2\right| &= \left|\mathbb{E}_{p^{\otimes n}}\left(T^n\right)^2 \chi_{t\ge a}\left(T^n\right)\right| \\ &= \left|\frac{1}{J^2}\mathbb{E}_{p^{\otimes n}}\left(\frac{1}{\sqrt{n}}L^{(n)}\right)^2 \chi_{t\ge\sqrt{n}Ja}\left(\frac{1}{\sqrt{n}}L^{(n)}\right)\right| \\ &\le \frac{1}{J^2}\frac{1}{n\left(Ja\right)^2}\mathbb{E}_{p^{\otimes n}}\left(\frac{1}{\sqrt{n}}L^{(n)}\right)^4 \\ &= \frac{1}{J^2}\frac{1}{n\left(Ja\right)^2}\left(\frac{n-1}{n}J + \frac{1}{n}\mathbb{E}_p\left(L\right)^4\right) = o\left(1\right). \end{aligned}$$

Similarly,

$$\begin{aligned} \left|\frac{1}{\sqrt{n}}\mathbb{E}_{p^{\otimes n}} L^{(n)} T_a^n - \frac{1}{\sqrt{n}}\mathbb{E}_{p^{\otimes n}} L^{(n)} T^n\right| &= \left|\frac{1}{\sqrt{n}}\mathbb{E}_{p^{\otimes n}} L^{(n)} T^n \chi_{t\ge a}\left(T^n\right)\right| \\ &= \left|\frac{1}{\sqrt{n}J}\mathbb{E}_{p^{\otimes n}}\left(\frac{1}{\sqrt{n}}L^{(n)}\right)^2 \chi_{t\ge\sqrt{n}Ja}\left(\frac{1}{\sqrt{n}}L^{(n)}\right)\right| = o\left(1\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{n} J_{q^n}\left(\delta'^n\right) &\ge \frac{\left|\frac{1}{\sqrt{n}}\mathbb{E}_{p^{\otimes n}} L^{(n)} T^n\right|^2}{\mathbb{E}_{p^{\otimes n}}\left(T^n\right)^2} + o\left(1\right) = \frac{1}{n}\frac{\left|\mathbb{E}_{p^{\otimes n}} L^{(n)} T^n\right|^2}{\mathbb{E}_{p^{\otimes n}}\left(T^n\right)^2} + o\left(1\right) \\ &= J_p\left(\delta\right) + o\left(1\right), \end{aligned}$$

which is (C0). $\blacksquare$

3.12 On asymptotic continuity

If for any $\{q^n, \delta'^n\}$ with $\|q^n - p^{\otimes n}\|_1 \to 0$ and $\frac{1}{\sqrt{n}} \left\|\delta'^n - \delta^{(n)}\right\|_1 \to 0$,

$$\varliminf_{n\to\infty} \frac{1}{n} \left| g_{q^n}\left(\delta'^n\right) - g_{p^{\otimes n}}\left(\delta^{(n)}\right)\right| = 0$$

holds, we say g is *asymptotically continuous* at $\left\{p^{\otimes n}, \delta^{(n)}\right\}$. Analogous conditions are used in study of entanglement measures etc. In our case, Fisher information satisfies '$\geq$', or weak asymptotic continuity, as stated in Theorem 13. However, the other side of inequality, and thus asymptotic continuity, is false. Let $p := \mathrm{Bin}\,(1,t)$, and

$$q^n\left(x^n\right) := \begin{cases} \frac{t}{2}^n, & \left(x^n = 0^n\right) \\ \left(1-t\right)^n + \frac{t}{2}^n, & \left(x^n = 1^n\right) \\ p^{\otimes n}\left(x^n\right), & \text{otherwise} \end{cases}$$

$$\delta'^n\left(x^n\right) := \delta^{(n)}\left(x^n\right),$$

$$\delta\left(0\right) = -\delta\left(1\right) = 1 > 0,$$

then we have $\|q^n - p^{\otimes n}\|_1 = \frac{t}{2}^n + \frac{t}{2}^n \to 0$, $\frac{1}{\sqrt{n}} \left\|\delta'^n - \delta^{(n)}\right\|_1 = 0$, and

$$\begin{aligned} \frac{1}{n} \left| J_{p^{\otimes n}}\left(\delta^{(n)}\right) - J_{q^n}\left(\delta'^n\right)\right| &= \frac{1}{n} \left| \left(\frac{1}{t^n} - \frac{2}{t^n}\right) n^2 + \left(\frac{1}{\left(1-t\right)^n} - \frac{1}{\left(1-t\right)^n + \frac{t}{2}^n}\right)\left(-n\right)^2 \right| \\ &= \frac{1}{n} \cdot n^2 \left| \frac{-1}{t^n} + \frac{\frac{t}{2}^n}{\left(1-t\right)^n \left\{\left(1-t\right)^n + \frac{t}{2}^n\right\}} \right| \to \infty. \end{aligned}$$

4 Classical Channels: Non-asymptotic theory

4.1 Axioms

Other than being square of a norm, $G_\Phi\left(\Delta\right)$ should satisfy:

(M1) *(monotonicity 1)* $G_\Phi\left(\Delta\right) \geq G_{\Phi\circ\Psi}\left(\Delta\circ\Psi\right)$

(M2) *(monotonicity 2)* $G_\Phi\left(\Delta\right) \geq G_{\Psi\circ\Phi}\left(\Psi\circ\Delta\right)$

(E) $G_{\Phi\otimes\mathbf{I}}\left(\Delta\otimes\mathbf{I}\right) = G_\Phi\left(\Delta\right)$

(N) $G_p\left(\delta\right) = J_p\left(\delta\right)$

4.2 Simulation of channel families

Suppose we have to fabricate a channel Φ_θ, which is drawn from a family $\{\Phi_\theta\}$, without knowing the value of θ but with a probability distribution q_θ or a channel Ψ_θ, drawn from a family $\{q_\theta\}$ or $\{\Psi_\theta\}$. More specifically, we need a channel Λ with

$$\Phi_\theta = \Lambda \circ (\mathbf{I} \otimes q_\theta)\,, \tag{14}$$

Here, note that Λ should not vary with the parameter θ. Giving the value of θ with infinite precision corresponds to the case where q_θ is delta distribution centered at θ.

Differentiating the both ends of (14) and letting $\Phi_\theta = \Phi$ and $q_\theta = q$, we obtain

$$\Delta = \Lambda \circ (\mathbf{I} \otimes \delta')\,, \tag{15}$$

where $\Delta \in \mathcal{T}_\Phi(\mathcal{C})$ and $\delta' \in \mathcal{T}_q(\mathcal{P}')$.

In the manuscript, we consider *tangent simulation*, or the operations satisfying (14) and (15), at the point $\Phi_\theta = \Phi$ only. Note that simulation of $\{\Phi, \Delta\}$ is equivalent to the one of the channel family $\{\Phi_{\theta+t} = \Phi + t\Delta\}_t$.

4.3 Relation between J and G

In this section, we review quickly the properties of norms with (M1), (M2), (E), and (N). For the proof, see [11].

Theorem 14 *Suppose (M1) and (N) hold. Then,*

$$G_\Phi(\Delta) \geq G_\Phi^{\min}(\Delta) := \sup_{p \in \mathcal{P}_{\mathrm{in}}} J_{\Phi(p)}(\Delta(p)) = \sup_{x \in \otimes_{\mathrm{in}}} J_{\Phi(\cdot|x)}(\Delta(\cdot|x))\,.$$

Trivially, $G_\Phi^{\min}(\Delta)$ satisfies (M1), (M2), (E), and (N).

Theorem 15 *Suppose (M2), (E) and (N) hold. Then*

$$G_\Phi(\Delta) \leq G_\Phi^{\max}(\Delta) := \inf_{\Lambda, q, \delta} \{J_q(\delta)\,;\ \Lambda \circ (\mathbf{I} \otimes q) = \Phi,\, \Lambda \circ (\mathbf{I} \otimes \delta) = \Delta\ \}\,.$$

Also, $G_\Phi^{\max}(\Delta)$ satisfies (M1), (M2), (E), and (N).

Obviously, $G_\Phi^{\min}(\Delta)$ and $G_\Phi^{\max}(\Delta)$ are not induced from any metric, i.e., they cannot be written as $S(\Delta, \Delta)$, where S is a positive real bilinear form. Indeed, we can show the following :

Theorem 16 *Suppose (M1), (N), and (E) hold. Then, $G_\Phi(\Delta)$ cannot written as $S_\Phi(\Delta, \Delta)$, where S is a positive bilinear form.*

5 Classical Channels: Asymptotic Theory

5.1 Asymptotic Theory: additional axioms

(A) *(asymptotic weak additivity)* $\lim_{n\to\infty}\frac{1}{n}G_{\Phi^{\otimes n}}\left(\Delta^{(n)}\right)=G_{\Phi}\left(\Delta\right)$

(C) *(weak asymptotic continuity)* If $\left\|\Phi^{n}-\Phi^{\otimes n}\right\|_{\mathrm{cb}}\to 0$ and $\frac{1}{\sqrt{n}}\left\|\Delta^{n}-\Delta^{(n)}\right\|_{\mathrm{cb}}\to 0$ then

$$\varliminf_{n\to\infty}\frac{1}{n}\left(G_{\Phi^{n}}\left(\Delta^{n}\right)-G_{\Phi^{\otimes n}}\left(\Delta^{(n)}\right)\right)\geq 0.$$

5.2 Asymptotic tangent simulation:definition

We consider asymptotic version of approximate version of (14)-(**??**):

$$\lim_{n\to\infty}\left\|\Phi^{\otimes n}\left(p\right)-\Lambda^{n}\left(p\otimes q^{n}\right)\right\|_{\mathrm{cb}}=0\,,\,\forall p, \tag{16}$$

and

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\left\|\Delta^{(n)}\left(p\right)-\Lambda^{n}\left(p\otimes\delta^{n}\right)\right\|_{\mathrm{cb}}=0,\,\forall p\,, \tag{17}$$

with "program" $\{q^{n},\,\delta^{n}\}$. Here, the larger one of $\left\|\Phi^{\otimes n}\left(p\right)-\Lambda^{n}\left(p\otimes q^{n}\right)\right\|_{\mathrm{cb}}$ and $\frac{1}{\sqrt{n}}\left\|\Delta^{(n)}\left(p\right)-\Lambda^{n}\left(p\otimes\delta^{n}\right)\right\|_{\mathrm{cb}}$ is called the *error* of the asymtotic tangent simulation.

5.3 Finite inputs

In this sutbsection, $\Omega_{\mathrm{in}}=\{1,\cdots,k\}$.

Theorem 17 *Suppose $\{\Phi\left(\cdot|x\right),\Delta\left(\cdot|x\right)\}$ satisfies all the conditions imposed on $\{p,\delta\}$ in Theorem 13. Let us define $\{q_{\varepsilon}^{n},\delta_{\varepsilon}^{n}\}:=\{\mathrm{N}\left(0,1\right),\delta\mathrm{N}\left(0,1\right)\}^{\otimes n(1+k\varepsilon)(J+c)}$ where $J=G_{\Phi}^{\min}\left(\Delta\right)=\max_{1\leq x\leq k}J_{\Phi(\cdot|x)}\left(\Delta\left(\cdot|x\right)\right)$ and $\varepsilon>0$, $c>0$ are arbitrary. Then, there is Λ^{n} such that*

$$\begin{aligned}\left\|\Phi^{\otimes n}\left(p\right)-\Lambda^{n}\left(p\otimes q_{\varepsilon}^{n}\right)\right\|_{\mathrm{cb}}&\leq\frac{C}{\left(\varepsilon n\right)^{1/4}},\\ \frac{1}{\sqrt{n}}\left\|\Delta^{(n)}\left(p\right)-\Lambda^{n}\left(p\otimes\delta_{\varepsilon}^{n}\right)\right\|_{\mathrm{cb}}&\leq\frac{C}{\left(\varepsilon n\right)^{1/4}},\end{aligned} \tag{18}$$

where C is a function of $\{\Phi\left(y|x\right),\Delta\left(y|x\right);x\in\Omega_{\mathrm{in}},y\in\Omega_{\mathrm{out}}\}$. Especially, if $|\Omega_{\mathrm{out}}|<\infty$, this function is countinuous and bounded.

Proof. Given the input sequence $x^{n}=x_{1}\cdots x_{n}$, denote the number of x in x^{n} by N_{x}. Suppose $N_{x}\geq\varepsilon n$. Then, we use $\{\mathrm{N}\left(0,1\right),\delta\mathrm{N}\left(0,1\right)\}^{\otimes N_{x}(J+c)}$ for simulation of $\{\Phi\left(\cdot|x\right),\Delta\left(\cdot|x\right)\}^{\otimes N_{x}}$. On the other hand, if

$N_x < \varepsilon n$, we first fabricate $\{\Phi(\cdot|x), \Delta(\cdot|x)\}^{\otimes \varepsilon n}$ using $\{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n\varepsilon(J+c)}$, and takes marginal. We do this for all $x = 1, \cdots, k$. Since

$$\bigotimes_{x=1}^{k} \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes N_x(1+\varepsilon)(J_x+c)} \equiv \{\mathrm{N}(0,1), \delta\mathrm{N}(0,1)\}^{\otimes n(1+k\varepsilon)(J+c)},$$

by Theorem 11 and Theorem 12, we have (18) and the proof is complete. ■

Theorem 18 *Suppose $\{\Phi(\cdot|x), \Delta(\cdot|x)\}$ satisfies all the conditions imposed on $\{p, \delta\}$ in Theorem 13 for all $x \in \Omega_{\mathrm{in}}$. Then, if a metric G satisfies (M1), (M2), (E), (A), (C), and (N0). Then,*

$$G_\Phi(\Delta) = G_\Phi^{\min}(\Delta).$$

Proof. Due to Theorem 14, we only have to show $G_\Phi(\Delta) \le G_\Phi^{\min}(\Delta)$. Consider the simulation of $\{\Phi, \Delta\}$ by $\{q_\varepsilon^n, \delta_\varepsilon^n\}$ as of Theorem 17. Due to Theorem 13,

$$G_{q_\varepsilon^n}(\delta_\varepsilon^n) = J_{q_\varepsilon^n}(\delta_\varepsilon^n).$$

Therefore, due to (18), we have

$$\begin{aligned}
0 &\underset{(\mathrm{C})}{\le} \varliminf_{n\to\infty} \frac{1}{n}\left(G_{\Lambda\circ(\mathbf{I}\otimes q_\varepsilon^n)}(\Lambda\circ(\mathbf{I}\otimes\delta_\varepsilon^n)) - G_{\Phi^{\otimes n}}\left(\Delta^{(n)}\right)\right) \\
&\underset{(\mathrm{M})}{\le} \varliminf_{n\to\infty} \frac{1}{n}\left(G_{\mathbf{I}\otimes q_\varepsilon^n}(\mathbf{I}\otimes\delta_\varepsilon^n) - G_{\Phi^{\otimes n}}\left(\Delta^{(n)}\right)\right) \\
&\underset{(\mathrm{E})}{=} \varliminf_{n\to\infty} \frac{1}{n}\left(G_{q_\varepsilon^n}(\delta_\varepsilon^n) - G_{\Phi^{\otimes n}}\left(\Delta^{(n)}\right)\right) \\
&= \varliminf_{n\to\infty} \frac{1}{n}\left(J_{q_\varepsilon^n}(\delta_\varepsilon^n) - G_{\Phi^{\otimes n}}\left(\Delta^{(n)}\right)\right) \\
&\le (1+\varepsilon k)\left(G_\Phi^{\min}(\Delta) + c\right) - \varlimsup_{n\to\infty} \frac{1}{n} G_{\Phi^{\otimes n}}\left(\Delta^{(n)}\right) \\
&\underset{(\mathrm{A})}{=} (1+\varepsilon k)\left(G_\Phi^{\min}(\Delta) + c\right) - G_\Phi(\Delta).
\end{aligned}$$

Since $\varepsilon > 0$ and $c > 0$ are arbitrary, we have

$$G_\Phi(\Delta) \le G_\Phi^{\min}(\Delta).$$

Finally, we show $G_\Phi^{\min}(\Delta)$ satisfies (C). Let $x_* \in \Omega_{\mathrm{in}}$ with $J_{\Phi(\cdot|x)}(\Delta(\cdot|x_*)) = G_\Phi^{\min}(\Delta)$, and $x_*^n = x_* x_* \cdots x_*$. Then, since $G_{\Psi^n}^{\min}(\Delta'^n) \ge J_{\Psi^n(\cdot|x_*^n)}(\Delta'^n(\cdot|x_*^n))$, we have

$$\varliminf_{n\to\infty} \frac{1}{n}\left(G_{\Psi^n}^{\min}(\Delta'^n) - G_{\Phi^{\otimes n}}^{\min}\left(\Delta^{(n)}\right)\right) \ge \varliminf_{n\to\infty} \frac{1}{n}\left\{J_{\Psi^n(\cdot|x)}(\Delta'^n(\cdot|x_*^n)) - J_{\Phi(\cdot|x)^{\otimes n}}\left(\Delta(\cdot|x_*)^{(n)}\right)\right\}.$$

The LHS of this is non-negative due to Theorem 13, since

$$\begin{aligned}
\left\|\Psi^n(\cdot|x_*^n) - \Phi(\cdot|x_*)^{\otimes n}\right\|_1 &\le \left\|\Psi^n - \Phi^{\otimes n}\right\|_{\mathrm{cb}} = o(1), \\
\frac{1}{\sqrt{n}}\left\|\Delta'^n(\cdot|x_*^n) - \Delta(\cdot|x_*)^{(n)}\right\|_1 &\le \frac{1}{\sqrt{n}}\left\|\Delta'^n - \Delta^{(n)}\right\|_{\mathrm{cb}} = o(1).
\end{aligned}$$

■

5.4 Continuous inputs

In this subsection, $\Omega_{\rm in}$ is a compact set in $\mathbb{R}^d$. Also, $\|x\|$ is usual 2-norm.

Theorem 19 *Suppose* $|\Omega_{\rm out}|<\infty$ *and*

$$\max\left\{\left\|\Phi\left(\cdot|x\right)-\Phi\left(\cdot|x'\right)\right\|_1,\left\|\Delta\left(\cdot|x\right)-\Delta\left(\cdot|x'\right)\right\|\right\}\leq f\left(\|x-x'\|\right)$$

holds for some $\lim_{t\to0}f\left(t\right)=0$. *Let us define* $\left\{q_\varepsilon^n,\delta_\varepsilon^n\right\}:=\left\{\mathrm{N}\left(0,1\right),\delta\mathrm{N}\left(0,1\right)\right\}^{\otimes n(1+k\varepsilon)(J+c)}$ *where* $J=G_\Phi^{\min}\left(\Delta\right)=\max_{1\leq x\leq k}J_{\Phi(\cdot|x)}\left(\Delta\left(\cdot|x\right)\right)$ *and* $\varepsilon>0$, $c>0$ *are arbitrary. Then, there is a family* $\left\{\Phi_t,\Delta_t\right\}_{t\geq0}$ *and* $\left\{\Lambda_t^n,q_{\varepsilon,t}^n,\delta_{\varepsilon,t}^{\prime n}\right\}_{t\geq0}$ *such that*

$$\left\|\Phi_t^{\otimes n}\left(p\right)-\Lambda_t^n\left(p\otimes q_{\varepsilon,t}^n\right)\right\|_{\rm cb}\leq\frac{C_t}{\sqrt{\varepsilon n}} \tag{19}$$

$$\frac{1}{\sqrt{n}}\left\|\Delta_t^{(n)}\left(p\right)-\Lambda_t^n\left(p\otimes\delta_{\varepsilon,t}^{\prime n}\right)\right\|_{\rm cb}\leq\frac{C_t}{\left(\varepsilon n\right)^{1/4}}\quad(\lim_{t\to0}C_t<\infty), \tag{20}$$

$$\lim_{t\to0}\left\|\Phi_t-\Phi\right\|_{\rm cb}=\lim_{t\to0}\left\|\Delta_t-\Delta\right\|_{\rm cb}=0. \tag{21}$$

Proof. Let $\mathcal{A}_t\subset\Omega_{\rm in}=\mathbb{R}^d$ be the totality of lattice points such that $\min_{x,y\in\mathcal{A}_t}\|x-y\|=t$. Define

$$\Phi_t\left(\cdot|x\right):=\Phi\left(\cdot|y\right),\quad\Delta_t\left(\cdot|x\right):=\Delta\left(\cdot|y\right),$$

where y is the closest point in $\mathcal{A}_t$ to x. By assumption, $\left\{\Phi_t,\Delta_t\right\}$ satisfies (21). By Theorem 17, we can compose Λ_t^n with (19) and (20). ■

(C2) If $\lim_{t\to0}\left\|\Phi_t-\Phi\right\|_{\rm cb}=\lim_{t\to0}\left\|\Delta_t-\Delta\right\|_{\rm cb}=0$, $\lim_{t\to0}G_{\Phi_t}\left(\Delta_t\right)=G_\Phi\left(\Delta\right).$

Theorem 20 *Suppose* $\left\{\Phi,\Delta\right\}$ *satisfies all the asumptions of Theorem 19. Then, if* G *satisfies (M1), (M2), (E), (A), (C), (N0), and (C2),*

$$G_\Phi\left(\Delta\right)=G_\Phi^{\min}\left(\Delta\right).$$

Proof. Again, we only have to show $G_\Phi\left(\Delta\right)\leq G_\Phi^{\min}\left(\Delta\right).$ By Theorem 19, $G_{\Phi_t}\left(\Delta_t\right)=G_{\Phi_t}^{\min}\left(\Delta_t\right)$. Therefore, due to (C2),

$$G_\Phi\left(\Delta\right)=\lim_{t\to0}G_{\Phi_t}\left(\Delta_t\right)=\lim_{t\to0}G_{\Phi_t}^{\min}\left(\Delta_t\right).$$

On the other hand, by construction of $\left\{\Phi_t,\Delta_t\right\}$,

$$G_{\Phi_t}^{\min}\left(\Delta_t\right)\leq\sup_x J_{\Phi(\cdot|x)}\left(\Delta\left(\cdot|x\right)\right)=G_\Phi^{\min}\left(\Delta\right).$$

Hence, we have the assertion. ■

5.5 Quantum states as a classical channel

A quantum state can be viewed as a channel which takes a measurement as an input, and outputs measurement result. Hence, if we restrict the measurements to separable measurements, the asymptotic theory discussed in this paper is applicable to quantum states also, proving the uniqueness of the metric. On the other hand, there are variety of monotone metrics, and lower asymptotic continuity is proven for some of them, e.g., SLD and RLD metric. This appearent contradiction can be circumvented by recalling that the theory of this paper is not applicable to the case of collective measurement.

References

[1] S. Amari, Differential-geometrical methods in statistics, Lecture Notes in Statistics, 28 (1985).

[2] S. Amari and H. Nagaoka, Methods of Information Geometry,Translations of Mathematical

[3] T. Cacoullos, and V. Papathanasiou, Characterizations of distributions by variance bounds. Statist. Probab. Lett. 7, 351-356 (1989). Monograph,Vol.191(AM Sand Oxford University Press, 2000).

[4] L. Le Cam and G. L. Yang, "Asymptotics in Statistics: Some Basic Concepts," (Springer, 2000).

[5] N. N. Cencov: Statistical Decision Rules and Optimal Inference. Trans. of Mathematical Monographs 53, Amer. Math. Soc., Providence (1982)

[6] A. Fujiwara, A Geometrical Study in Quantum Information Systems, doctoral thesis (1995).

[7] L. Goldstein, and G. Reinert, Stein's Method and the Zero Bias Transformation with Application to Simple Random Sampling, Annals of Applied Probability, 7, 935-952 (1997).

[8] L. Goldstein, and Y. Rinott,.On multivariate normal approximations by Stein's method and size bias couplings, J. Appl. Prob. 33, 1-17 (1996).

[9] K. Matsumoto, A Geometrical Approach to Quantum Estimation Theory, doctoral dissertation, University of Tokyo, 1998.

[10] K. Matsumoto, Reverse estimation theory, Complementarity between RLD and SLD, and monotone distances, arXiv:quant-ph/0511170 (2005).

[11] K. Matsumoto, On metric of classical channel spaces : non-asymptotic theory, arXiv:quant-ph/1001.3547(2010).

[12] K. Matsumoto, in preparation

[13] T. Mikami, Equivalent conditions on the central limit theorem for a sequence of probability measures on R. Statist. Probab. Lett. 37, 237-242 (1998).

[14] Cacoullos, T., Papadatos, N. and Papathanasiou, V., "Three Elementary Proofs of the Central Limit Theorem with Applications to Random Sums " in: Stochastic Processes and Related Topics (I. Karatzas, B.S. Rajput and M.S. Taqqu, Eds.), 1998, Birkhauser, Boston, pp. 15-23.

[15] D. Petz, Monotone metrics on matrix spaces, Linear Algebra Appl., 244,81–96 (1996).

[16] H. Strasser, "Mathematical theory of statistics," (Walter de Gruyter, 1981).

[17] A. N. Shiriaev and V. G. Spokoiny, "Statistical decision theory and experiments: asymptotic theory," (World Scientific, 2000).

[18] E. Torgersen, "Comparison of statistical experiments," (Cambridge university press ,1991).